Chapter 4
Single Stage Produce-to-Stock

Better customer service
Lower manufacturing costs (setups)
More uniform utilization
Target-Level Production Authorization (PA) Mechanism with Backlogs

\[ I(t)B(t) = 0 \]

\[ C(t) + I(t) = Z \text{ tags} \]
Cumulative Counts & Relationships

Up to time $t$,

$A(t) = \text{number of customer arrivals}$

$D(t) = \text{number of items produced}$

$R(t) = \text{number of items delivered to customers}$

If $Z$ items were in the store at time 0,

$R(t) = \min\{Z + D(t), A(t)\}$

$B(t) = A(t) - R(t)$

$I(t) = Z + D(t) - R(t)$

$C(t) = \min\{A(t) - D(t), Z\}$
Fictitious Queuing System

Assume $I(0) = Z$, $C(0) = 0$.

$A_n$ is the time of the $n$th customer arrival

$D_n$ is the time when the $n$th item is received at the store

$S_n$ is the time to produce the $n$th item

Then $D_{n+1} = \max \{D_n, A_{n+1}\} + S_{n+1}$

is exactly like the departure process from a single-server queue that begins empty at time 0 and has arrival times $A_n$ and service times $S_n$.

Let $N(t)$ be the number of customers in this fictitious system.
Thinking About the Fictitious System

• If a customer arrives when \( I(t) > 0 \)
  – Customer gets item and tag is sent to factory.
  – Arrival of customer to store = arrival of tag to factory

• If a customer arrives when \( I(t) = 0 \)
  – Customer waits until an item is completed, at which time a tag is sent to the factory.
  – Queue of backordered customers can be seen as being appended to the queue of \( Z - 1 \) tags at the factory (1 tag in service); when an item is completed, the first backordered customer leaves with that item and is replaced in the factory queue by the newly completed item’s tag. Factory queue has no size limit.
  – Arrival of customer at store = arrival of entity to factory
Using the Fictitious System

\[ N(t) = A(t) - D(t) \] is the number of jobs in the fictitious system.

We can find all the quantities in the real system in terms of \( N(t) \):

\[ I(t) - B(t) = Z + D(t) - A(t) = Z - N(t) \]

\[ I(t) = \max \{ Z - N(t), 0 \} \]

\[ B(t) = \max \{ N(t) - Z, 0 \} \]

\[ C(t) = \min \{ N(t), Z \} \]

\[ B(t) + C(t) = N(t) \]

So “all” we need is the distribution of \( N(t) \)!
Poisson Arrivals, Expon. Processing Times: Fictitious System is M/M/1

\[ P\{N = n\} = \rho^n(1 - \rho), \quad n = 0, 1, 2, \ldots \quad (\rho < 1) \]

So \( I = \max\{Z - N, 0\} \) implies that

\[ P\{I = 0\} = P\{N \geq Z\} = (1 - \rho)\sum_{n=Z}^{\infty} \rho^n = \rho^Z \quad \text{(backorder)} \]

and for \( n = 1, \ldots, Z \),

\[ P\{I = n\} = P\{N = Z - n\} = (1 - \rho)\rho^{Z-n} \]

Similarly, \( B = \max\{N - Z, 0\} \) implies that

\[ P\{B = 0\} = P\{N \leq Z\} = (1 - \rho)\sum_{n=0}^{Z} \rho^n = 1 - \rho^{Z+1} \]

and for \( n = 1, 2, \ldots \)

\[ P\{B = n\} = P\{N = n + Z\} = (1 - \rho)\rho^{n+Z} \]
M/M/1 Fictitious System (cont.)

Also $C = \min\{N, Z\}$ implies that for $n = 0, 1, \ldots, Z-1$,

$$P\{C = n\} = P\{N = n\} = (1 - \rho)\rho^n$$

and $P\{C = Z\} = P\{I = 0\}$ or

$$P\{C = Z\} = P\{N \geq Z\} = (1 - \rho)\sum_{n=Z}^{\infty} \rho^n = \rho^Z$$

From these distributions we can find expected values:

$$E[B] = \frac{\rho^{Z+1}}{1 - \rho}$$  \hspace{1cm} \text{backorders}$$

$$E[C] = \frac{\rho}{1 - \rho} \left(1 - \rho^Z\right)$$  \hspace{1cm} \text{jobs in factory}$$

$$E[I] = Z - E[C] = Z - \frac{\rho}{1 - \rho} \left(1 - \rho^Z\right)$$  \hspace{1cm} \text{stock in store}
Choosing Z

The purpose of producing to stock is to improve customer service (reduce the wait to receive an item). But this goal must be balanced against the cost of carrying a store of inventory.

Suppose $k_1$ is the inventory carrying cost ($/item/unit time) and $k_2$ is the backlogging cost ($/item/unit time). Then

$$\Delta TC(Z) = TC(Z) - TC(Z - 1) = k_1 - (k_1 + k_2) \rho^Z$$

is convex in $Z$. The incremental cost of the $Z^{th}$ item in inventory is

$$TC(Z) = k_1 E[I] + k_2 E[B]$$
Choosing Z (cont.)

Convexity means that as $Z$ increases, $\Delta TC(Z)$ changes from negative to positive. Find the $Z$ where it is closest to 0 by

$$k_1 - (k_1 + k_2) \rho^Z = 0$$

$$Z^* = \frac{\log\left(\frac{k_1}{k_1 + k_2}\right)}{\log(\rho)}$$

Then round up or down to the closest integer.

Other considerations are that the probability a customer is backlogged is $\rho^Z$ and the expected time to fill a demand is $\rho^2/(\mu - \lambda)$. Note that without the store ($Z = 0$ if we produce to order) these quantities are 1 and $1/(\mu - \lambda)$, respectively.
Target-Level Production Authorization (PA) Mechanism with Lost Sales

\[ C(t) + I(t) = Z \text{ tags} \]
Corresponding Queuing System

- If a customer arrives when $I(t) > 0$
  - Customer gets item and tag is sent to factory.
  - Arrival of customer to store = arrival of tag to factory
- If a customer arrives when $I(t) = 0$
  - Customer leaves again without an item, no tag is sent to the factory.
  - Potential arrival of tag to factory is blocked if $Z$ tags are already in the factory
- Behaves like $GI/G/1/Z$: $Z$ is a limit on the number of jobs in the system.
Using the Corresponding System

Assume $I(0) = Z$

$N(t) = R(t) - D(t)$ is the number of jobs in the limited-population system

We can find all the quantities in the real system in terms of $N(t)$:

$\dot{I}(t) = Z - N(t)$

$C(t) = N(t)$

So “all” we need is the distribution of $N(t)$!
Poisson Arrivals, Expon. Processing Times: Fictitious System is M/M/1/Z

Note: This is different from the finite source (machine interference) model in that the population limit is imposed on the system and is not just a consequence of having a finite number of potential customers!!
Corresponding $M/M/1/Z$ System

$$p(n) = P\{N = n\} = \frac{(1-\rho)\rho^n}{1-\rho^{Z+1}}, n = 0,1,...,Z$$

So $I = Z - N$ implies that for $n = 0, 1, \ldots, Z$,

$$P\{I = n\} = P\{N = Z - n\} = \frac{(1-\rho)\rho^{Z-n}}{1-\rho^{Z+1}}$$

$$E[I] = \frac{(1-\rho)(Z + \rho^{Z+1}) - \rho(1-\rho^{Z+1})}{(1-\rho)(1-\rho^{Z+1})}$$

And $C = N$ implies that

$$P\{C = n\} = P\{N = n\} = \frac{(1-\rho)\rho^n}{1-\rho^{Z+1}}, n = 0,1,...,Z$$

$$E[C] = \frac{\rho(1-\rho^{Z+1}) - (1-\rho)(Z+1)\rho^{Z+1}}{(1-\rho)(1-\rho^{Z+1})}$$
Choosing $Z$

Service level is the proportion of customers who obtain products: \( SL = P\{I > 0\} = 1 - p(Z) \)

Suppose \( k \) is the inventory carrying cost ($/item/unit time) and \( v \) is the profit ($/item). Then the total profit per unit time (to be maximized) is

\[
TP(Z) = v \lambda (1 - p(Z)) - k E[I]
\]
Multiple Product Target-Level PA Mechanism with Backlogs

Item types $i = 1, \ldots, r$

$$C_i(t) + I_i(t) = Z_i \text{ tags for type } i$$

Customers for type $i$: rate $\lambda_i$

Item types $i = 1, \ldots, r$

$$I_i(t)B_i(t) = 0$$

$$I_i(t) > 0 \quad I_i(t) = 0$$

Store $I_i(t)$

$B_i(t)$

rate $\mu$

m/c

tag

item + tag
Fictitious Multiclass M/M/1 System

- If a type $i$ order arrives when $I_i(t) > 0$
  - Customer gets item and type $i$ tag is sent to factory.
  - Arrival of order for type $i$ = arrival of type $i$ “customer” to M/M/1
- If a type $i$ order arrives when $I_i(t) = 0$
  - Customer waits until a type $i$ item is completed, at which time a type $i$ tag is sent to the factory.
  - Queue of type $i$ backorders can be seen as being appended to the queue of $Z_i$ tags at the factory (including possibly 1 tag in service); when a type $i$ item is completed, the first type $i$ backorder is filled and that item’s tag is returned to the factory. Factory queue has no size limit.
  - Arrival of order for type $i$ = arrival of type $i$ “customer” to M/M/1
Multiclass M/M/1 Steady State Dist’n

Population vector \( \mathbf{N}(t) = (N_1(t), N_2(t), ..., N_r(t)) \)

is not a Markov chain because type of item in process affects the transition probabilities;

but the total population \( N(t) = \sum_{i=1}^{r} N_r(t) \) is CTMC with

\[
P\{N = n\} = \rho^n(1 - \rho), \quad n = 0, 1, 2, \ldots, \text{ where}
\]

\[
\rho_i = \frac{\lambda_i}{\mu} \quad \text{and} \quad \rho = \sum_{i=1}^{r} \rho_i
\]

Use “multinomial thinning” to get

\[
\Pr [ N_1 = n_1, ..., N_r = n_r ] = \frac{n!}{n_1!n_2!...n_r!} (1 - \rho)^{n_1} \rho_1^{n_1} \cdots \rho_r^{n_r}
\]
Marginal Steady State Dist’n

Then the marginal distribution of the number of type \( i \) customers in the M/M/1 system is

\[
\Pr[N_i = n_i] = (1 - \hat{\rho}_i) \hat{\lambda}_i^{n_i}, \quad n_i = 0, 1, \ldots \text{ where } \hat{\rho}_i = \frac{\lambda_i}{\mu - \sum_{j \neq i} \lambda_j}
\]

So \( I_i = \max\{Z - N_i, 0\} \) implies that

\[
P\{I_i = 0\} = P\{N_i \geq Z_i\} = (1 - \hat{\rho}_i) \sum_{n=Z_i}^{\infty} \hat{\rho}_i^n = \hat{\rho}_i^{Z_i}
\]

and for \( n_i = 1, \ldots, Z_i \),

\[
P\{I_i = n_i\} = P\{N_i = Z_i - n_i\} = (1 - \hat{\rho}_i) \hat{\rho}_i^{Z_i-n_i}
\]

Similarly, \( B_i = \max\{N_i - Z_i, 0\} \) implies that

\[
P\{B_i = 0\} = P\{N_i \leq Z_i\} = (1 - \hat{\rho}_i) \sum_{n=0}^{Z_i} \hat{\rho}_i^n = 1 - \hat{\rho}_i^{Z_i+1}
\]

and for \( n_i = 1, 2, \ldots \),

\[
P\{B_i = n_i\} = P\{N_i = n_i + Z_i\} = (1 - \hat{\rho}_i) \hat{\rho}_i^{n_i+Z_i}
\]

Find \( Z_i \) in the same way we found \( Z \) for a single product type.
Multiple Product Target-Level PA Mechanism with Lost Sales

Item types $i = 1, \ldots, r$

$$C_i(t) + I_i(t) = Z_i \text{ tags for type } i$$

Store

$B_i(t) = 0$

$I_i(t) = 0$

$I_i(t) > 0$

Customers for type $i$: rate $\lambda_i$

Item

Rate $\mu$

m/c

Item + tag

tag
Corresponding Queuing System

• If a type $i$ order arrives when $I_i(t) > 0$
  – Customer gets item and type $i$ tag is sent to factory.
  – Arrival of order for type $i = $ arrival of type $i$ “customer”

• If a type $i$ order arrives when $I_i(t) = 0$
  – Customer leaves again without an item, no tag is sent to the factory.
  – Potential arrival of tag to factory is blocked if $Z_i$ tags are already in the factory

• Behaves like $GI/G/1/Z$: $Z_i$ is a limit on the number of type $i$ jobs in the system.
Poisson Arrivals, Same Expon.
Processing Time Dist’n for Each Type

Corresponding queuing system is M/M/1/Z

\[ p(n) = \lim_{t \to \infty} \Pr[N(t) = n] \]

\[ = G(Z)^{-1} \frac{n!}{n_1! \cdots n_r!} \rho_1^{n_1} \cdots \rho_r^{n_r} , \quad 0 \leq n_i \leq Z_i, \quad i = 1, \ldots, r, \]

where \( n = n_1 + \ldots + n_r \) and

\[ G(Z) = \sum_{n_1=0}^{Z_1} \cdots \sum_{n_r=0}^{Z_r} \frac{(n_1 + \ldots + n_r)!}{n_1! \cdots n_r!} \rho_1^{n_1} \cdots \rho_r^{n_r} \quad \text{normalizing constant} \]

service level

\[ SL_i = 1 - \Pr[I_i = 0] = \frac{G(Z - e_i)}{G(Z)} \quad e_i \text{ is } r\text{-vector of 0's with 1 in the } i\text{th place} \]
Generalized PA Systems

# of Production Authorization Cards ≠ Z
Relationship Between Tags and Physical Inventory

\( I(t) = \) Physical inventory in output store
\( B(t) = \) Backlog of unmet demands (requisition tags); \( I(t)B(t) = 0 \)
\( Z = \) Limit on physical inventory
\( K^+(t) = \) # of process tags in the output store
\( K^-(t) = \) # of order tags waiting for match with process tags; \( K^+(t)K^-(t) = 0 \)
\( K = \) Limit on number of process tags

\[ Z - I(t) + B(t) = K - K^+(t) + K^-(t) \]

is the number of items to produce if no arrivals after \( t \)
Time Lag Between Orders and Requisitions

- Order tag can arrive before requisition tag
  - As advance notice from the customer, or
  - If forecast of future orders is accurate.
- Assume
  - There is a fixed time lag, \( \tau \), between receipt of an order tag and receipt of the corresponding requisition tag, i.e., if the \( n^{th} \) order tag arrives at time \( A_n \), then the \( n^{th} \) requisition tag arrives at time \( A_n + \tau \)
  - At time 0, there are \( Z \) items and \( K \) PA cards in the output store (factory idle)
- We need to keep track of
  - Processes related to the order tags, PA cards and process tags
  - Processes related to the requisition tags, physical inventory, and backlogged demand.
Order tags, PA cards, Process tags

Up to time $t$,

$A(t) =$ the number of order tags that have arrived,
$D(t) =$ the number of (items + process tags) sent to the output store
$O(t) =$ the number of PA cards sent to factory

At time $t$,

$C(t) =$ the number of PA cards in the factory
Order tags, PA cards, Process tags (cont-1)

At time $t$,

$$O(t) = \min \{K + D(t), A(t)\}$$

$$C(t) = O(t) - D(t) = \min \{K, A(t) - D(t)\}$$

$$K^+(t) = K - C(t) = K - D(t) + O(t)$$

$$K^-(t) = A(t) - O(t)$$

Define: $N(t) = A(t) - D(t)$

is population of a queuing system with arrivals corresponding to arrivals of order tags and departures corresponding to transfers of completed items to the output store.
Order tags, PA cards, Process tags (cont-2)

Then  \[ K^+(t) - K^-(t) = K + D(t) - A(t) = K - N(t) \]

\[ K^+(t) = \max\{K - N(t), 0\} \]

\[ K^-(t) = \max\{N(t) - K, 0\} \]

\[ C(t) = \min\{K, N(t)\} \]

\[ C(t) + K^-(t) = N(t) \]

Then \(K^+(t), K^-(t), C(t)\) can be found in terms of the “fictitious” queuing system as we found \(I(t), B(t),\) and \(C(t)\) previously in the ordinary target-level PA system with backlogging.
Requisition Tags, Inventory, Backlogs

Up to time $t$,

$$R(t) = \text{the number of requisition tags that have arrived},$$

$$S(t) = \text{the number of items delivered to customers}$$

At time $t$,

$$I(t) = Z + D(t) - S(t)$$

$$S(t) = \min \{Z + D(t), R(t)\}$$

$$B(t) = R(t) - S(t)$$
Requisition Tags, Inventory, Backlogs (cont.)

Define: \[ N^*(t) = R(t) - D(t) \]

\[ I(t) - B(t) = Z - N^*(t) \]

\[ I(t) = \max \{ Z - N^*(t), 0 \} \]

\[ B(t) = \max \{ N^*(t) - Z, 0 \} \]

Thus if we can find the distribution of \( N^*(t) \), then we can find the distributions of \( I(t) \) and \( B(t) \) as before.
Relationship Between $N(t)$ and $N^*(t)$

Since there is a fixed time lag, $\tau$, between receipt of an order tag and receipt of the corresponding requisition tag,

$$R(t) = A(t - \tau)$$

Then

$$N^*(t) = R(t) - D(t) = A(t - \tau) - D(t)$$

$$= A(t - \tau) - D(t - \tau) + D(t - \tau) - D(t)$$

$$= N^*(t) - (\text{the number of departures from } N(t)\text{'s queuing system between } t - \tau \text{ and } t).$$

If orders follow a Poisson process, service times are exponential and there is a single server, then $N(t)$ is M/M/1 and

$$P\left\{N^*(t) = n\right\} = e^{-\mu \tau (1 - \rho)} P\left\{N(t) = n\right\} = e^{-\mu \tau (1 - \rho)} \rho^n (1 - \rho)$$
Choosing $Z$ and $\tau$

From the distribution of $N^*(t)$, we can find

$$E[B] = e^{-\mu \tau (1-\rho)} \frac{\rho^{Z+1}}{1-\rho}$$

$$E[I] = Z + \lambda \tau - \frac{\rho}{1-\rho} (1 - \rho^Z e^{-\mu \tau (1-\rho)})$$

and

$$TC(Z, \tau) = k_1 E[I] + k_2 E[B]$$

Further analysis shows that to minimize $TC$, it is optimal to set $Z^* = 0$ and

$$\tau^* = -\ln\left(\frac{k_1}{k_1 + k_2}\right) E[T]$$

where $E[T]$ is the mean flow time if the system were operated as produce-to-order; also $TC(Z^*, \tau^*)$ is lower than the minimum total cost without advance orders.